Section 6:
Using Assessment to Improve Program Results – Understanding Children's Growth, Family Experiences, and Program Effectiveness

Deborah (Montgomery) Parrish, Managing Research Scientist
American Institutes for Research

Introduction

With increasing state and national interest in the potential of preschool programs to prepare young children for kindergarten and beyond, there has been a corresponding rise in the call to document child outcomes and program results for policymakers, program administrators, and families. This emphasis on results-based accountability is part of a growing movement to link information about how children, families, and programs are doing (i.e., their outcomes or results) with deliberate strategies for ongoing program improvements that enhance these results. From its inception, the First 5 California Commission on Children and Families has been committed to results-based accountability. This section focuses on the appropriate uses of assessment to understand preschool children's growth and development, families' experiences with preschool programs, and the implementation of evidence-based practices that contribute most to the achievement of identified program results for children and families.

The section begins with an overview of child-, family-, and program-based assessments that are applicable to preschool, followed with greater detail about the appropriate uses of each to inform program improvements that enhance child and family results. Next, the section shows how the three types of assessment are integrated in California's Desired Results System. The section also provides examples of how localities in California are using assessment to measure children's developmental progress and to improve the quality of programs. Finally, the section offers some practical suggestions for communities on how to evaluate preschool programs responsibly, using multiple measures to confirm findings and identify trends.

Overview of Types of Assessment

There are three major types of assessment – child-based, family-based, and program-based – applicable to measuring the quality and impact of preschool programs. Each is suited to unique purposes and program goals. A clear set of goals and objectives for using these types of assessment is therefore the first step in planning how assessment tools are to be selected, implemented, and reported.

Child-based assessments for young children include:

- Developmental screening to identify the need for more-in-depth assessment of possible disabilities or other special needs;
- Diagnostic tests to identify and confirm specific disabilities or other special needs;[3]
- Developmental profiles or observations to aid teachers in documenting children's developmental progress in their natural daily environments for the purpose of improving curriculum and instruction to enhance results; and
- Specially created exercises performed by the child in the presence of a trained assessor to provide an objective snapshot of children's performance that, when aggregated for groups

[3] Diagnostic assessments are designed to be used only by specially trained professionals. This section does not include a discussion of assessments used only for diagnostic purposes.

of children, can assess the effectiveness of prekindergarten programs in achieving desired outcomes (also called "direct assessments").

Standardized paper-and-pencil tests completed independently by the student are not recommended for use with preschool children. These types of child-based assessments may, however, be used appropriately in third or fourth grade in elementary school as part of the longitudinal tracking of the effectiveness of preschool programs.

Family-based assessments, including written surveys, personal interviews, focus groups and other feedback sessions may be used to:

- Determine how well the program is meeting child and family needs;
- Assess family members' progress in achieving family or personal goals;
- Obtain suggestions for program improvement; and
- Explore how best to engage families as partners in their children's learning.

Program-based assessments include environment rating tools and other types of quality and compliance checklists or procedures (e.g., focus groups of teaching staff to get a sense of what is working and what could be improved) for:

- Periodic self-study by program staff;
- Monitoring by state agencies and other program sponsors; and
- Research by outside evaluators.

Taken together, the information gained from these various types of assessments—about children, families, and programs—can be used to inform immediate, short-term, and long-term goals for planning and program quality improvement. For example, suppose developmental profile results for children as a group show they are making good progress with social development (as appropriate for their age), but are not progressing in the area of early literacy (e.g., letter knowledge, word knowledge, phonemic awareness, concepts of print, and story comprehension). Suppose also that family assessments reveal that families express a desire for more information about early literacy activities that they can use at home. Finally, suppose that the program self-study using the environmental rating scale reveals that the program is either "inadequate" or "minimal" in its ratings on the language-reasoning subscales (Books and pictures, Encouraging children to communicate, Using language to develop reasoning skills, and Informal use of language). The program would then have multiple sources of evidence indicating that its early literacy activities need to be strengthened. The thesis presented here is that assessment is most informative when multiple measures are used in this way to confirm findings or trends that can be identified for targeted program improvements that enhance child and family results.

The following pages provide greater detail about the uses of child-based, family-based, and program-based assessments for program quality improvement. Examples from state-funded school readiness and child development preschool programs provide a context for how local programs are using assessment to inform implementation and ongoing quality improvement to achieve results for children and families.

Child-based Assessment: Its Uses (and Potential Abuses) for Preschool Children

There are a number of questions that must be addressed when thinking about implementing a child assessment system for preschool children:

- For what purpose will the child assessment information be used?
- What outcomes should be measured?
- What instruments should be used?
- Who should provide input regarding the child's developmental status? Can data be collected from the child, the parent, and/or or the teacher?
- When and how often should the assessments be administered and by whom?
- Should formal or informal methods be applied?
- Should standardized measures be used?
- Is the instrument adequately comprehensive in its coverage of key developmental domains, yet user-friendly to teachers and families?
- Can the assessment instrument be integrated with developmentally appropriate curriculum and naturally occurring typical daily activities?
- What kinds of risks and benefits are involved?
- How will the data be analyzed to inform practice?
- What is the intended use of the data?
- What safeguards should be put in place to protect against unintended use?

Defining the purposes of assessing young children

First and foremost, it is important to articulate the purpose (or purposes) for which young children are to be assessed. As we have seen, the intended purpose directly influences the design and implementation of the assessment system. In short, the intended use determines how the assessment is carried out (i.e., frequency, duration, and type of assessor training needed), what kind of assessment takes place (i.e., specific domains of learning and development assessed), and the means of data collection (i.e., naturalistic observation-based versus direct assessment).

In 1998, the National Education Goals Panel (NEGP) put forth the *Principles and Recommendations for Early Childhood Assessments* (Shepard, Kagan and Wurtz, 1998), in which they listed general principles to guide policy and practice, and identified four broad purposes for which early childhood assessment is currently used:

1) To promote learning and development of individual children,
2) To identify children with special needs and health conditions for intervention purposes,
3) To monitor trends in programs and evaluate program effectiveness,
4) To obtain benchmark data for accountability purposes at the local, state and national level.

Clarification of the primary purpose(s) for which early childhood assessments are to be used is not always straightforward. Different stakeholders (e.g., teachers, parents, administrators,

program evaluators, and policymakers) will each have their own view of why children should or should not be assessed in particular ways. Teachers will be interested in assessment tools that inform their daily practice with individual children or identify children with special needs. Parents want clear and useful information about how their child is progressing in multiple domains of learning. Program administrators may be more interested in group measures of how children of a particular age or within a particular classroom are doing, while program evaluators, monitoring agencies, or policymakers may be most interested in implementing an accountability system that provides appropriate forms of documentation that programs are having the intended effects over time, across the curriculum, and for the population of enrolled students as a whole.

In particular, program planners must be careful not to attempt to target multiple purposes of early childhood assessment with a single tool, unless clear safeguards are in place that ensure that individual child measures are not used for high-stakes decisionmaking, holding a child back from attending kindergarten, or otherwise inappropriately labeling a child on the basis of a single assessment. Equally important, a single assessment measure is not sufficient for accountability purposes, such as determining whether the funding of a program should be expanded or withdrawn. As suggested above, assessment affecting young children, or the programs that serve them, is most reliable when it involves multiple measures, each of which provides findings pointing in the same general direction.

Table 6-1, which is excerpted from the NEGP report, shows how the appropriate uses and technical accuracy of various types of assessment change with the age of child being assessed and the purpose for which the assessments are to be used. As children grow older, direct measures and the aggregation of child results may be appropriate for high-stakes uses and monitoring purposes. However, the panel recommended that high-stakes assessments intended for accountability purposes (e.g., high stakes decisions about individual children and programs) should not be used for children until they are at least eight years of age (the end of third grade or preferably fourth grade). Consideration of the types of assessment tools to be selected must begin with consensus among stakeholders as to the specific purpose(s) for which the assessments will be used as well as the degree of technical accuracy that is appropriate to expect for different ages of children. Particular care must be taken to avoid the use of direct or indirect assessment measures to make high-stakes decisions about individual preschool children.

Appropriate Uses and Technical Accuracy of Assessments Change Across the Early Childhood Age Continuum (Birth to Age 8)

Birth 1 2 3 4	Kindergarten 5	1st grade 6	2nd grade 7	3rd grade 8 years	Beyond age 8
Purpose 1: Assessing to promote children's learning and development					
Parents and caregivers observe and respond as children develop language and physical skills.	Parents, caregivers, and preschool teachers use direct measures, including observations of what children are learning, to decide what to teach next.	Teachers use both formal and informal assessments to plan and guide instruction.			
Purpose 2: Identifying children for health and special services					
All children should be screened regularly for health needs, including hearing and vision checks, as part of routine health care services. Many serious cognitive and physical disabilities are evident at birth or soon thereafter. As soon as developmental delays or potential disabilities are suspected, parents and physicians should seek in-depth assessments.	Children entering Head Start and other preschool programs should be screened for health needs, including hearing and vision checks. Individual children with possible developmental delays should be referred for in-depth assessment.	All children should be screened at school entry for vision and hearing needs and checked for immunizations. Some mild disabilities may only become apparent in the school context. Districts and states must by law have sound teacher and parent referral policies, so that children with potential disabilities are referred for in-depth assessment.			
Purpose 3: Monitoring trends and evaluating programs and services					
Because direct measures of children's language and cognitive functioning are difficult to aggregate accurately for ages from birth to 2, state reporting systems should focus on living and social conditions that affect learning and the adequacy of services.	Assessments, including direct and indirect measures of children's physical, social, emotional, and cognitive development, could be constructed and used to evaluate prekindergarten programs, but such measures would not be accurate enough to make high-stakes decisions about individual children.	Beginning at age 5, it is possible to use direct measures, including measures of children's early learning, as part of a comprehensive early childhood assessment for monitoring trends. Matrix sampling should be used to ensure technical accuracy and to provide safeguards for individual children. Because of the cost of such an assessment, states or the nation should pick one grade level for monitoring trends in early childhood, most likely kindergarten or first grade.			
Purpose 4: Assessing academic achievement to hold individual students, teachers, and schools accountable					
				Before age 8, standardized achievement measures are not sufficiently accurate to be used for high-stakes decisions about individual children and schools. Therefore, high-stakes assessments intended for accountability purposes should be delayed until the end of third grade (or preferably fourth grade).	

Table 6-1. (Source: Shepard, Kagan and Wurtz, 1998)

What are the current issues and concerns surrounding assessment of young children?

With new scientific knowledge about how children develop, increased awareness of the importance of fostering school readiness through high quality preschool programs, and the rise in accountability pressures across the country, the demand for valid and reliable early childhood assessment and performance measures has risen markedly since the NEGP report was issued. This trend is marked by the requirement to establish outcome measures for all Head Start programs, and by the rise in the number of state-funded preschool programs with mandatory child assessment systems (Horton and Bowman, 2002).

New mandates for direct skill-based assessments of all children in Head Start programs have spurred heated discussions about the risks and dangers of assessment-based high-stakes decisions. This is especially true if there are questions about the cultural and linguistic appropriateness of such assessments for the population with whom they are to be used. Early childhood experts argue that it is difficult to use standardized testing methods to obtain valid and reliable results with young children—they are generally considered to be "poor test takers." In addition, early childhood is a period when "children's rates of physical, motor, and linguistic development outpace growth rates at all other stages. Growth is rapid, episodic, and highly influenced by environmental supports: nurturing parents, quality caregiving, and the learning setting" (Shepard, Kagan, and Wurtz, 1998). Thus, experts caution against the use of assessment results to rank, sort, or retain young children, or to draw hasty conclusions about program quality and funding. There is a real danger of misclassifying young children, particularly English learners or children with special needs, potentially causing them to miss out on the most optimal learning opportunities if tracked into inappropriate learning environments.

In designing an assessment system, it is also important to distinguish between knowledge and learning ability when considering what is appropriate to assess. Children may enter a program having different learning backgrounds, knowledge, and skills, but are often quick to learn and may grow at different rates across the major learning domains (social-emotional, cognitive, physical, and adaptive) and during different phases of the early years. Hence, tracking and labeling based on a single assessment at a point in time should be avoided, particularly in early childhood. The use of assessments to sort and track children can also result in a wider gap between those labeled as "ready" and "unready" for school (Shepard, Kagan, and Wurtz, 1998). These gaps can be exacerbated if practitioners resort to placements of children into homogeneous settings or ability-based groupings based on assessment results. On the surface, such placements may appear to be easy solutions for teachers dealing with children across a wide range of abilities, but research confirms that these arrangements do not promote optimal learning environments for young children. To ensure that assessments meet the intended goal of benefiting every child, every assessment system should include preventive measures that safeguard against misuse.

Risks Associated With Early Childhood Assessment (Adapted from Muenchow, 2003)

1) **The misuse of assessment data can result in major consequences for individual children.** This includes the use of data to draw inappropriate conclusions, such as the use of test results to deny children Kindergarten entrance. Young children are difficult to assess and preschool assessment results are neither adequate nor reliable enough to justify holding them back from kindergarten.

2) **Assessment tools that do not distinguish between knowledge and learning ability have the potential to inappropriately underestimate children's development** by not accounting for the diverse learning backgrounds, knowledge, and skills that children have, but that may not be measured by traditional knowledge- or skill-based assessment instruments. Assessments should be culturally sensitive—respecting the diversity of cultures, languages, and special needs of the children for whom they are used.

3) **Assessment activities entail a diversion of scarce resources** from program expansion or other quality improvement activities. Consideration must be given to the degree to which teachers' time and energy is diverted from other responsibilities in order to conduct assessments. The burden and costs of assessments should not outweigh their benefits. Assessment systems benefit teachers when they are clearly linked to curriculum and instruction decisions.

4) **Programs serving the most disadvantaged children could be penalized.** If assessment outcomes are used to determine funding allocations, programs that need the most support may not receive it. Funding policies should not create disincentives for programs to serve (or recruit) children with special needs or English language learners. Diverting public resources away from children who could most benefit from services results in higher service costs over the long-term.

Minimizing Risks and Maximizing Benefits of Early Childhood Assessment

(Adapted from Muenchow, 2003)

1) **Include child development specialists and diverse stakeholders in the design and implementation** of the assessment system.

2) **Develop guidelines for the use of assessment data** to benefit children and families.

3) **Be clear about the major purpose(s) for developing an assessment system,** recognizing that "one size does not fit all." Create an awareness that a single instrument cannot address all of the purposes of early childhood assessment.

4) **Consider the costs of the assessment system.** This includes the cost of the instrument as well as the cost of administering the assessment, teacher training, and data management and reporting. It is also important to consider which agency will be paying for it.

5) **Take a family-centered and inclusive approach.** Assessments of young children must take into account the considerable influences of the family system and home environment on their development. Elicit input from families and use a universal design that is appropriate for *all* children, including English learners and those with special needs.

6) **Release the assessment data in aggregate form, as part of a larger data set that addresses multifaceted aspects** of the program (e.g., program quality and family background of participants). This will help to guard against simplistic interpretations of the assessment data.

7) **Select assessment tools that are technically sound: check for validity and reliability.** Validity is the accuracy of the tool in measuring what it is intended to measure. Reliability is the degree to which the assessment measures consistently across different instances of measurement—such as across raters, times of measurement, or sets of items.

Consideration of the various agents that will be using the data gathered from the assessments is also key. To the extent that program administrators, county officials and state officials may have conflicting interests (or goals), the assessment system may generate biased outcomes depending on the incentives or disincentives the system includes. For example, if a county is required to submit data to the state for a statewide program rating, data collected by county officials may tend to overestimate the progress made by children enrolled in their programs. This is likely to be true in cases where children's collective progress in a program is associated with the amount of state funding it can secure. In such instances, not only may the assessment system generate false results, it may also penalize programs that are committed to serving children with more challenging needs.

Studies have shown that ratings of children tend to be accurate among early childhood teachers who have had appropriate training, but that biases may be introduced when their ratings are known to influence program funding (Meisels & Atkins-Burnett, 2000). This type of punitive approach to funding is *not* the case in California, where state funds for child development programs are apportioned based on established formulas, and programs are not held accountable for specific benchmarks of child achievement. Rather, state-funded programs are responsible for demonstrating that individual children are making progress over time and that programs are implementing program quality standards and are in compliance with all contract requirements.

Guiding principles for the assessment of young children

The National Education Goals Panel developed the following major themes in their guidelines (Shepard, Kagan, & Wurtz, 1998) regarding the design and practice of assessment in early childhood. These themes are in agreement with more recent recommendations in a position statement from the National Association for the Education of Young Children in collaboration with the National Association of Early Childhood Specialists in State Departments of Education (2003):

(1) First and foremost, the assessment should be beneficial to the child and helpful to teachers in modifying the curriculum and instruction to better meet the needs of the child.

The assessment should be age-appropriate in content and include all developmental and learning domains pertaining to young children. These include the social, emotional, cognitive, physical, and adaptive domains, the latter being particularly important for children with special needs.[4] The content of the assessment should be relevant to program goals and integrated with the curriculum, so that teachers can use assessment information to guide their day-to-day instructional practices.

[4] Domains as identified by a U.S. Office of Special Education Program-sponsored expert work group on early childhood outcomes, January 2003.

(2) Administrators and practitioners should use age-appropriate methods of data collection.

Instead of paper-and-pencil tests that generally are neither adequate nor appropriate for assessing young children, the assessment should be based on ongoing observations across a range of naturally occurring activities that take place as part of the child's everyday routine. In addition, information about a child's growth and development should be gathered and documented at regular intervals, over an extended period of time. This is necessary since observations at any one point in time do not provide a complete picture of a child's development. Multiple methods should be used to gather information, including input from all of the teachers who work with the child as well as family members, anecdotal records, photographs, examples of children's work and behaviors, language samples, and other documentation of children's developmental status. In general, the assessment system as a whole should emphasize a child's strengths and competencies, and foster a child's confidence and desire for learning.

(3) It is also important to acknowledge diversity and individual differences in the design and implementation of assessment measures.

More specifically, assessment procedures and instruments should accommodate differences in cultural and linguistic background, as well as a variety of learning styles and learning rates. Keep in mind that all assessments, to a certain extent, are measures of one's language ability (Shepard, Kagan, Wurtz, 1998). In other words, assessment results are easily confounded by language proficiency. Each child's first and second language development should be taken into account when determining appropriate assessment methods and interpreting the meaning of assessment results. Similarly, the concept of "cultural competence," or "effectively operating in different cultural contexts" is relevant to assessment systems, in that assessments should be fair and nondiscriminatory for children from diverse backgrounds or with special needs (Cross, et. al, 1989).

(4) Finally, it is critical to include parents, teachers and other adult caregivers in a collaborative process of child assessment.

The emphasis on naturalistic observation in the above guidelines is supported by the recommendations of national leaders in the field of early childhood development. In a survey conducted by the Erikson Institute, experts in the assessment of young children identified a strong match between the program curriculum and the skills to be assessed as one of the most important characteristics of a sound assessment system (Horton and Bowman, 2002).

With regard to assessment data, the most favored methods included the use of teacher observations and meetings (e.g., weekly teacher meetings to examine individual cases and to improve the curriculum) and the use of portfolios (i.e., the collection of student work samples). The least favorable methods included standardized tests and work sheets. The majority of the experts surveyed viewed standardized tests as a poor means of assessing young children. A smaller group suggested that standardized tests could be used for program evaluation (by way of pre/post design carried out on a district-wide sample), but cautioned against the use of such instruments for individualized assessments of each child. The study also surveyed state-funded

preschool programs to examine child assessment practices across the nation. Almost 70 percent of state-funded preschool programs reported that they mandate or widely use observational techniques in their assessment systems. This is consistent with the recommendations of the early childhood experts who were surveyed.

Two additional guidelines are included here as important considerations with regard to assessing children in California's inclusive and linguistically diverse preschool programs:

(5) Utilizing a "universal design" in the development and implementation of assessment instruments ensures that measures are used appropriately for *all* children.

Universal design implies that all measures have been reviewed and revised to include language that applies to the broadest population possible. For example, language such as "child points to musical toy" might be written as "child attends to musical toy", so that children with vision impairments or physical limitations have the best possibility of demonstrating their performance, based on the intent of the measure (in this case, a measure to demonstrate that the child is aware of a particular object in his or her environment).

The federal Individuals with Disabilities Education Act (IDEA) requires that states establish performance goals for students with disabilities that are consistent, to the maximum extent appropriate, with other goals and standards in place for non-disabled students. In addition, the law requires that states include children with disabilities in the general state and district-wide assessment programs (i.e., those intended for typically developing children), with appropriate accommodations, as necessary. Further, states are required to develop guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in state and district-wide assessment programs. Thus, children with disabilities are to be assessed using the same measures as are used for typically developing children, unless it can be shown that alternate assessments are more appropriate for particular individuals. A universal design helps to ensure that the broadest range of children with special needs are included in district- or state-level assessment systems for the general population.

(6) Assessments of English learners should be administered by persons who understand the child's primary means of communication.

As described earlier, assessment results are easily confounded by language proficiency. Preschool children who are English learners may take longer to produce language because they are absorbing the sounds, vocabulary and conventions of two or more languages. Thus, since early childhood assessments rely to some extent on children's language abilities to assess their development in other domains (cognitive, social, etc.), English learners may be inappropriately penalized if the assessor does not speak or understand the child's primary language. If the child's teacher does not communicate fluently in the child's primary language, the assistance of someone who does should be sought in order to document observations, determine the child's level(s) of development, and track child progress with accuracy.

General Principles to Guide Policy and Practice in the Assessment of Preschoolers

Principles and Recommendations for Early Childhood Assessments (Shepard, Kagan, & Wurtz, 1998) lists six general principles for guiding policymakers and practitioners in their design of assessments for young children. These include:

- Assessment should bring about benefits for children—either in direct services or improved quality of educational programs.
- Assessments should be tailored to a specific purpose and should be reliable, valid, and fair for that purpose.
- Assessment policies should be designed recognizing that reliability and validity of assessments increase with children's age.
- Assessments should be age-appropriate in both content (i.e., addressing the full range of early learning and developmental domains) and the method of data collection (i.e., in familiar contexts and without reliance on paper-and-pencil tasks).
- Assessments should be linguistically and culturally appropriate, recognizing that to some extent, all assessments are measures of language.
- Parents should be a valued source of assessment information, as well as an audience for assessment results. Assessments should include multiple sources of evidence, especially reports from parents and teachers. Sharing results with parents should be part of an ongoing process that involves parents in their child's education.

Screening children for possible special needs

Screening children for possible special needs is an essential function of child assessment. The purpose of this type of screening is not to make a diagnosis, but rather to identify children in need of referral for a more in-depth assessment of special needs. Research suggests that early identification of special needs, and early intervention to address them, has many benefits. Children with disabilities who receive early-intervention services show "significant" developmental improvements even after only one year of service, according to a report to Congress by the federal Department of Education (2003). Moreover, there is evidence that the earlier the identification of special needs and the onset of intervention, the better. Hence, ideally, many disabilities and special needs will be identified and treated long before a child reaches preschool age. However, since some types of special needs are not easily recognized prior to the preschool years, screening at the preschool level offers a critical opportunity to identify children with special needs.

The First 5 California Commission on Children and Families' new California Special Needs Project will include, among other deliverables, the development of a universal safeguards protocol to (1) identify and/or develop the tools and strategies for screening all children; (2) to develop trainings and train trainers and implementers; and 3) to work with the contract evaluator to assess the effectiveness of screening tools and strategies. Hence, this section will not go into depth on the topic of screening children for possible special needs. We offer here only the following general guidelines:

- Upon entry to preschool programs, all children should be screened for health needs, including hearing and vision checks.
- Early screening through the periodic use of parental report screening instruments should be encouraged. The American Academy of Sciences has identified valid, reliable, sensitive and specific tools for identifying children in need of further assessment. These include the Parents' Evaluation of Developmental Status (PEDS),[5] the Ages and Stages Questionnaires,[6] and the Child Development Inventories[7] as appropriate parental screening tools for preschool-age children (Dunkle, M. & Vismara, L., 2003).
- Since screening measures are, by design, quick, shortened versions of more in-depth assessments, they are generally less reliable. Thus, they should not be used as a substitute for comprehensive in-depth assessment. The potential benefits of early intervention are great, but so is the possible harm that could result from inappropriate labeling and treatments resulting from inadequate or inaccurate assessments. Screening measures should never be the only assessment used to identify children for special education.
- Screening tools should be used to identify and refer children who appear to need more in-depth assessments. Referrals to the appropriate specialist(s) should be made for those children whose screening results suggest the need for more comprehensive evaluation.
- Teachers, parents, and physicians should seek in-depth assessments as soon as developmental delays or potential disabilities are suspected.
- For potential cognitive or language disorders, measures should meet the highest standards of reliability and validity, be administered and interpreted by trained professionals, include multiple sources of evidence in both home and school settings, and be used in conjunction with primary language assessments for children with more than one language (i.e., English learners) (Shepard, Kagan and Wurtz, 1998).

Assessing Family Experiences and Satisfaction

Effective preschool programs not only support children's growth and development, but also support their families. Family members are children's first teachers, and research suggests that partnerships with families have substantial long-term benefits for children (Henderson & Berla, 1994). Thus, assessing families' experiences with preschool programs is another important form of preschool-related assessment.

Although "family-centered" programs are increasingly seen as the most effective approach to enhancing the development of young children, researchers and evaluators have only recently begun to come to consensus on the identification of specific family-related benefits or "outcomes" to be expected from preschool programs. There has been relatively less discussion regarding how best to measure family outcomes. Whereas program accountability regarding interactions with families generally rests with ensuring that staff members provide families with information about the program and its policies, their child's developmental status and progress, and opportunities for involvement and parent conferences, many programs are incorporating the

[5] For more information, see http://www.pedstest.com/test/peds_intro.html
[6] For more information, see http://www.brookespublishing.com/store/books/bricker-asq/index.htm
[7] For more information, see http://www.firstsigns.org/downloads/CDIs_subdoc.PDF

concept of "family partnership" as a key program goal. Family-staff partnerships, or relationships based on trust, mutual respect, open communication, and a collaborative attitude, are acknowledged as integral to the provision of family-centered services. Research also confirms that family involvement in a child's education increases the child's long-term achievement and school success (Henderson and Berla, 1994).

Programs can use assessments of family experiences and satisfaction to find out whether families perceive staff to be responsive to families' needs, to welcome family members as partners in their children's learning and development, and to encourage family involvement in meaningful ways. In addition, input from parents can help programs identify information that parents want, or provide an opportunity for parents to express more general or specific suggestions for program improvement. Evaluation of program strategies to promote effective two-way communication between staff and families is another way that assessment can be used for program improvement. Increasingly, programs are striving to move beyond general measures of family satisfaction toward assessments that provide a more comprehensive understanding of family perspectives regarding program efforts to foster family partnerships that support children's learning and development.

Measures of family satisfaction, although important, can be difficult to interpret, since family members may not have a standard against which to judge the services their child is receiving (Simeonsson, 1988). In addition, studies suggest that most parents report a high degree of satisfaction with services their children receive, so satisfaction measures may not be as sensitive to changes in programs and services over time (McNaughton, 1994). Nonetheless, parent satisfaction with services is a critical outcome, since it provides a check on the degree to which programs are providing services that are family-centered and satisfying to their users. In addition, consumer satisfaction has been related to more active participation and follow-through in medical and educational services, as well as to the perceived benefits of services (Bailey, et.al. 1998).

To avoid some of the pitfalls of more general family satisfaction measures, the design of a family assessment tool should ensure that the measures cover the range of services provided (e.g., information exchange, communication styles and formats, development and attainment of child and family goals) yet are specific enough to differentiate between various program practices and their impacts on families and their children (Bailey et al., p. 317).

While few child development staff question the importance of partnerships with families, many talk about a variety of challenges associated with building and maintaining partnerships with families. Challenges include long work hours for parents, cultural or linguistic barriers, lack of understanding on the part of parents and/or staff about the importance of parent involvement, and a lack of training among child development staff to communicate and engage effectively with family members. Despite the many challenges facing families, national survey data indicate that participating in their children's education is a priority among families, regardless of their education or socio-economic status.

Assessment of program policies and procedures that impact communication and collaboration with families is one way to identify training needs and areas for improvement. A number of

program self-assessment tools have been identified by the CDE in its *Family Partnership Initiative Training Manual.* Since effective partnerships begin with good two-way communication, assessment of staff communication skills from the perspective of both staff and family members is a good place to begin. A basic element of good communication with families is the ability of staff to understand and be responsive to each child's family and cultural traditions and practices. Programs that succeed in establishing effective two-way communication with families can then take steps to strengthen their partnerships with families in numerous other ways. They are more able to respond to families' expressed needs and suggestions, and are open to offering the types of opportunities that maximize family members' involvement in and support of their child's learning both in the program and at home. Only when family members have and take advantage of the opportunity, through parent conferences, advisory committees, or other outlets, to express their own vision of family participation in the life of the program, will a true partnership between staff and families begin to develop. Programs can provide such opportunities by first determining the communication formats that work best for the families served and then using these formats to encourage two-way exchanges via parent conferences, meetings, parent and child together times, take-home reading programs, family surveys, or other modes of communication.

Assessing Program Quality

Child and family assessments that inform programs about their direct impacts are useful for shaping day-to-day instructional activities and implementing family-friendly processes, but are insufficient for identifying program areas to target for program quality improvement. Assessments designed to evaluate children's learning environments are now widely used. They are a component of California's monitoring system for state-funded child development programs, and for evaluation of school readiness programs and early childhood staff training and retention incentive program quality.[8] These environmental rating tools can be used for periodic self-study by program staff, by outside evaluators for monitoring program quality, or by researchers as objective measures of program quality. For program staff, they provide a self-check on major program components related to program quality (e.g., space and furnishings, activities, adult-child interactions, and program structure), and for evaluators and researchers, they are used to assess the overall quality of preschool programs, to target technical assistance, or to measure the impact of program quality improvement efforts.

There are two types of program quality that child development researchers agree are important: *process* quality and *structural* quality. Process quality includes the interactions children have with adults, peers, and materials, whereas structural quality refers to the types of standards that may be subject to regulation, such as adult-child ratios, group size, and staff qualifications. Both impact children's learning and development; process quality does this most directly, and is

[8] Use of standardized environment rating scales for program self-study are now required for state-funded child development programs on a yearly basis. Their use is also required by a number of counties receiving First 5 California funds to implement compensation and retention, training, and school readiness programs.

influenced by aspects of structural quality.[9] Ongoing assessment of program quality is key to early identification of areas for improvement and maintenance of high quality standards.

The aspects of structural quality that consistently predict high process quality in early childhood programs include teacher education in early care and education, staff compensation, stability of teaching staff (i.e., low teacher turnover), adult/child ratios, and group size. Staff training and certification that lead to levels of compensation comparable to that of teachers working in the K-12 system and professional standards for adult/child ratios are two factors that are associated with the lowest levels of teacher turnover and improved child outcomes. Incorporating goals and benchmarks for these components of structural quality should be part of every preschool program's self-assessment process. California's licensing requirements for center- and family child care home-based programs constitute a foundation for the basic components of structural quality, such as health and safety, adult/child ratios, and staff qualifications. Program quality reviews to determine compliance with California's child development standards or eligibility for accreditation by the National Association for the Education of Young Children include components of both process and structural quality, and require periodic updates that are designed to maintain high standards of program quality over time.

In addition, a widely used standardized environment rating scale for preschool programs is the ***Early Childhood Environment Rating Scale (ECERS)*** (Harms, Clifford, and Cryer, 1998)**.** The ECERS-R (revised in 1998) primarily assesses process quality—the interactions between all members of the preschool classroom environment and the interactions children have with the materials in their environment—as well as some of the structural features such as space, schedule and materials that support these interactions. The ECERS tool relies on careful observation of the child's environment, such as the interactions between them and program staff or peers, interactions between them and classroom activities, supervision, arrangement of space, and other factors that support their learning. The scale does not measure structural indicators such as staff to child ratio and class size, but it does include some aspects of structural quality. In general, the scale contains items that examine the physical environment, curriculum, schedule and structure, basic care, interactions, discipline, and parent and staff education.
ECERS is built on the understanding that children have three basic and equally important needs: 1) the protection of their health and safety, 2) the nurturing of positive relationships with parents, family, community and other children, and 3) the opportunities for stimulation and experiential learning (Harms, Clifford and Cryer, 2000). The early childhood environment is defined broadly—the arrangement of indoor and outdoor space, the materials and activities offered to children, the supervision and interactions (including language) that occur, and the schedule of the day, including routines and activities. Support offered to parents and staff is also included. This tool is accompanied by a complete multimedia training program (including interactive video and workbook), and has been widely used in the assessment of preschool program quality within and outside the state of California, both for ongoing program monitoring and large-scale research projects. It is considered to be reliable and valid, and suitable for the evaluation of inclusive and culturally diverse programs.

[9] Superintendent's Universal Preschool Task Force Report, p. 18 (Kontos and others 1995; Whitebook and others 1990; Howes and others 1992)

Using the ECERS-R, a trained observer visits and observes a classroom for approximately three or four hours (or enough time to observe greetings and departures and all activities in between). The observer selects from among brief descriptions of tangible aspects of the environment that are arranged along a continuum ranging from "inadequate" to "minimal" to "good" to "excellent" (a numerical scale of from 1 to 7), for each of 43 subscales. Findings can be profiled on a single page, to highlight trends by major theme, and to pinpoint individual subscales in need of improvement. Because the descriptions are very concrete, it is easy for the user to know exactly what is needed to move from a subscale rating of "minimal" to "good," for example.

Although inter-rater reliability on the ECERS-R requires rigorous training and regular checks against independent "anchor" reliability raters[10] from outside of the program, the instrument is being used more broadly by program staff who are not necessarily trained to achieve this level of inter-rater reliability. In these cases, the ratings can be instructive for self-study and program improvement, but should not be relied upon for high stakes accountability (Whitebook, Sakai, Howes, and Wishard, 2003). The CDE/CDD now requires all state-funded preschool programs to conduct an annual program self-study and report on their findings (including an action plan based on key findings) using the ECERS-R. The CDE/CDD Field Services branch uses selected subscales as part of a comprehensive program review, and compares its findings with those of the program every three years.[11] In addition, training and use of the scales is a requirement for staff receiving stipend incentives from First 5 California-funded compensation and retention initiative programs that exist in most counties. Recently funded school readiness programs are also using the ECERS as a tool to assess prekindergarten program quality.

Another tool used by programs to evaluate program quality is the ***Assessment Profile for Early Childhood Programs***[12] (Abbott-Shim & Sibley, 1998), an observational checklist containing 147 Yes/No items designed to self-assess five aspects of the classroom environment: the learning environment, curriculum, interactions, individualizing, and health and safety. The *Assessment Profile* has been used in a number of national studies of early childhood programs (For example, the FACES battery includes the Learning Environment and Scheduling subscales of the ***Assessment Profile for Early Childhood Programs***).

A commonly used measure of process quality (teacher-child interactions) is the ***Caregiver Interaction Scale*** (Arnett, 1989) that rates teachers' sensitivity during their interactions with children. The measure consists of 26 items that yield three scores (sensitivity—warm, attentive, engaged; harshness—critical, punitive; detachment—low levels of interaction, interest, or supervision) which are combined to create an overall caregiver quality score. The ratings are made after two 45-minute observations conducted on two separate occasions by two separate observers. The ***Caregiver Interaction Scale*** is relatively easy to learn to use and it has been included in numerous studies of child care quality.

[10] To increase inter-rater reliability, assessors may be trained to become reliability "anchors" to ensure consistency of ratings among users of the ECERS-R for research or other purposes requiring this level of rigor.
[11] Contract Monitoring Reviews (CMR) are conducted every three years for non-LEA based programs, and Coordinated Compliance Reviews (CCR) are conducted every four years for LEA-based programs.
[12] For on-line information about the *Assessment Profile for Early Childhood Programs*, see http://www.qassist.com/assm.htm

A relatively new environmental assessment instrument that is less widely known but which focuses on an area of increasing state and national importance is the ***Early Language and Literacy Classroom Observation (ELLCO)*** (Smith and Dickinson, 2002). This is a field-tested observation tool, designed for use in prekindergarten through third grade classrooms to assess environmental factors that are specifically related to literacy and language development in young children. The toolkit includes an environment checklist, a classroom observation and teacher interview instrument, and a literacy activities rating scale. The assessment may be administered by program directors, supervisors, or teachers. It helps programs to assess their classrooms on 14 variables that span four functional areas: classroom functional environment, the interactive environment, language and literacy facilitation, and broad support for literacy.

These types of rating scales are also useful as teaching tools for staff, because their design makes very clear exactly what is needed to improve program quality in specific areas. Teachers have found them to be straightforward to use, meaningful in terms of identifying areas for improvement, and informative in assessing changes in program quality over time.

Putting It All Together: California's *Desired Results for Children and Families* System

Many states have developed, or are currently developing, early childhood standards and assessment systems in this era of results-based accountability.[13] California's *Desired Results for Children and Families* System has the merit of incorporating all three types of assessment described above—child-based assessment in the form of the Desired Results Developmental Profile (DRDP), family-based assessment in the form of an annual family survey, and program-based assessment in the form of annual use of the Early Childhood Environment Rating Scale (ECERS) described above, and a process for conducting ongoing program self-evaluation, which programs must all implement in order to receive state funding.

The *Desired Results* system is unique in that is was designed to include *all* children. Guidelines for environmental or other adaptations that can be used by teachers to assess children with disabilities on the same set of DRDP measures are incorporated as part of Desired Results *Access*—the parallel project implemented through collaboration between CDE's Child Development and Special Education Divisions.

Background on the Development of the Desired Results System

The state Department of Education, Child Development Division, developed the *Desired Results for Children and Families* system between 1996 and 2000 to improve the quality of the child development services it provides. The system is designed to document the progress made by children and families toward the achievement of desired results. By documenting progress over time, the Desired Results system provides concrete information to help practitioners and program administrators focus on specific program components, as necessary, to improve program quality.

[13] A recent survey by Project SPARC found that 30 states have developed early childhood outcome standards, but relatively few states have also developed inclusive assessment systems that are aligned to these standards.

The Desired Results for children include:

- Children are personally and socially competent.
- Children are effective learners.
- Children show physical and motor competence.
- Children are safe and healthy.

The Desired Results for families include:

- Families support their children's learning and development.
- Families achieve their goals.

The CDE/CDD acknowledges that there are many contributors to the achievement of these desired results. The most basic assumption behind this set of results is that no single program or type of program is capable of achieving the desired results by itself, and therefore no one program can be held solely accountable for that achievement or lack of achievement. The desired results are achieved by the combined contributions of the network of local, regional, and statewide early care and development services, conditioned, of course, by the larger environment within which children and families live.

At the same time, there is an assumption that each component within the network of services does in fact make its own contribution to the achievement of desired results. This assumption is already implicit in the multiplicity of programs that form the CDE child development service system. Each program was created for a specific purpose that expresses the manner in which that program is intended to contribute to the desired results. The measures included in the desired results structure are designed to capture the actual level of that contribution.

What is the Desired Results for Children and Families System?

The Desired Results System is a results-based accountability system that is aligned with the California Department of Education's overall goals for children and families and is linked to the language, literacy, and mathematics content and performance standards for kindergarten through 12th grade. The system was designed in 1996 by the California Department of Education, Child Development Division to do the following:

- Serve as a framework for documenting progress of all children and families
- Give teachers concrete information to use to modify curriculum
- Document how all children are benefiting from programs (for community- and state-level policymakers)
- Target technical assistance for ongoing program quality improvement

Progress of children and families is assessed through the structured observations of children (the Desired Results Developmental Profiles or "DRDPs") and input from parents (via family surveys and parent conferences). Evaluation of program quality is accomplished through the use of standardized Environment Rating Scales and compliance review instruments. Together, the information provided by the DRDPs, family surveys, and self-studies using the environment rating scales contribute to an overall assessment of program strengths and weaknesses. Program staff complete an "action plan" on a yearly basis that uses the findings about children's progress, family experiences and satisfaction, and program quality to determine areas to focus on and action steps for program improvements leading to enhanced child and family results.

The Desired Results Developmental Profile[14]

The Desired Results Developmental Profile (DRDP), a component of the *Desired Results for Children and Families* system described above, is California's response to the design of a child-based assessment. The Desired Results Developmental Profile (DRDP) is a structured observation tool that helps teachers to track children's progress over time across key domains of development, as described by the four desired results for children. The model uses a research-based conceptual framework that describes how children typically progress in key domains of development (e.g., social-emotional, cognitive, language and literacy, early math, physical and health). The DRDP thus provides teachers and caregivers with a framework for organizing their observations of children so that children's progress can be documented. This type of assessment is "curriculum-embedded," in that it is integrated with developmentally and age-appropriate instructional practices and thus can be used to inform curricular decisions for individual children and groups of children.

For each desired result (e.g., "Children are personally and socially competent"), there are from one to four *indicators*, which are clusters of related developmental themes that define the desired result more precisely so that it can be measured. The indicators describe the specific aspects of development that would indicate that a desired result is being achieved. The indicators for each of the four desired results for children are listed below:[15]

Desired Result 1: Children are personally and socially competent.

Indicators:
- **1-1.** Children show self-awareness and a positive self-concept.
- **1-2.** Children demonstrate effective social and interpersonal skills.
- **1-3.** Children demonstrate effective self-regulation in their behavior.
- **1-4.** Children show growing abilities in communication and language.

Desired Result 2: Children are effective learners.

Indicators:
- **2-1.** Children are interested in learning new things.
- **2-2.** Children show cognitive competence and problem-solving skills through play and daily activities.
- **2-3.** Children show interest in real-life mathematical concepts.
- **2-4.** Children demonstrate emerging literacy skills.

Desired Result 3: Children show physical and motor competence.

Indicator:
- **3-1:** Children demonstrate an increased proficiency in motor skills.

[14] The description of the DRDP is based on recent work by CDE to revise the instrument in order to improve its validity and reliability. The revised version is expected to be available to the field during the 2005-2006 school year.
[15] The Desired Results and Indicators span children's development from birth to five years (infants, toddlers, and preschoolers) and from five through 12 years for children who are served in before- and after-school programs.

Desired Result 4: Children are safe and healthy.

Indicator: 4-1: Children show an emerging awareness and practice of safe and healthy behavior.

The Desired Results and Indicators are broad statements that cannot be measured directly. Thus, for each indicator, there are two or more *developmental themes* that are the specific domains of development that can be measured over time. For example, the developmental themes for preschoolers that are included for Desired Result 1 (Children are personally and socially competent), and Indicator 1-1 (Children show self-awareness and positive self-concept), are "identifies self," "sense of own ability to do things," and "expression of self: feelings." These are the concepts for which three to five developmental levels are described from which teachers select the most appropriate one for the child they are assessing. The developmental levels (or "landmarks" of development) within each theme are described and illustrated using examples of children's behaviors or language that teachers can observe during typical daily routines in developmentally appropriate program activities. The examples depict the types of behaviors that would have to be observed in order for a child to demonstrate mastery of each developmental level. Teachers select the developmental level that best describes the behavior or skill that the child typically and regularly exhibits.

Using these frameworks, each DRDP describes a continuum of development in areas such as social skills, language, and motor skills. Instead of setting up artificial testing situations, teachers use naturalistic observation techniques, document their observations (i.e., anecdotal records), and select the developmental level (for each developmental domain and sub-domain, or "developmental theme") that best matches the developmental level of the child they are observing. Thus, teachers can map a child's progress along the model as he or she grows and develops throughout the year, and over the course of a number of years. There are four DRDP instruments, covering a continuum of development across the same set of desired results and indicators for children from birth through 12 years: birth to 3 years, 3 years through prekindergarten, 5 through 8 years, and 9 through 12 years.

The DRDP instruments have been mandated for use by CDD-funded programs since 2001-2002, which marked the beginning of a 4-year phased-in implementation period to include all state-funded center-based programs and family child care home networks by 2005-2006. At that time, and for the first time in the history of the CDE/CDD, all of these programs will be using the same set of results-based structured observations to track children's progress toward the achievement of desired results across a broad number of developmental areas. The DRDP Access instrument includes the same measures as the DRDPs, with the addition of guidance for environmental and other adaptations to be used (as appropriate) for children ages birth to five years who have Individual Education Programs (IEPs) or Individualized Family Service Plans (IFSPs).

The Desired Results system requires that teachers develop the observational skills necessary to use the DRDP appropriately. In addition, once teachers map an individual child's development using the DRDP, they are required to use this information for curricular planning, so that

children's development is supported and enhanced. Training will be needed on an ongoing basis to help teachers conduct meaningful observations and to use the information to support children's learning and development. CDE is also in the process of developing and refining electronic formats of the DRDP that will be used to generate summary information about individual children and groups of children, in addition to developing other reporting formats to aid programs in their use of the data. The system will also allow users to integrate their findings about children, families, and program quality with curriculum planning and program improvement strategies.

California's Desired Results Family Survey

To assess families' experiences and satisfaction with their child's preschool program, the California Department of Education has developed the *Desired Results Family Survey*. The survey is based upon the two desired results for families, namely that "families support their child's learning and development" and that "families achieve their goals."
Because programs are keenly aware of the limits of their influence in impacting *how* families support their children's learning, the survey focuses on those practices that are within a program's locus of control: providing information to families about a range of topics (and being responsive to families' expressed desires for specific types of information), including their child's developmental progress; providing a safe and nurturing environment for children so that family members' work and life goals can be achieved; and providing opportunities for family members' involvement and participation in the program.

The *Desired Results Family Survey* includes questions about families' satisfaction with the types of information they receive from the program, the ways in which program staff communicate with and support them and their child, and program characteristics, such as indoor and outdoor space, materials and equipment, activities, and staff.

In addition to surveys, family perspectives can be obtained via personal interviews, focus groups, feedback sessions following special events or meetings involving families, or by parent-run data gathering efforts. To protect family confidentiality, programs should be willing to accept anonymous input from families who do not wish to identify themselves.

CDE/CDD-funded child development programs are required to administer the *Family Survey* on an annual basis. Survey results are summarized and compared with information gleaned from the child and program quality assessments, in order to determine appropriate goals and action steps for program improvements that enhance child and family outcomes. In this way, program staff can look for corroborating evidence indicating specific areas in need of improvement. For example, if child data show that a group of four-year-old children are not showing progress in the area of conflict resolution with their peers, family survey data show that families are not satisfied with the ways in which children play and interact with each other, and program quality ratings show that the program is "inadequate" on the subscale for "interactions among children," program staff might conclude that they should consider new strategies to foster more positive interactions between children, and should take steps to communicate these strategies to family members so that they could reinforce them at home.

Integrating Assessment Data to Improve Program Results for Children and Families

Integrating aggregated assessments of children, family experiences, and program quality can provide meaningful information for evaluating program strengths and identifying areas for improvement to enhance child and family results. Program staff can begin to use these data sources to look for emerging themes, identify corroborating evidence or to determine areas for targeted follow-up data gathering. Training sessions for programs preparing to implement the Desired Results system in California now include practice in using these types of data for program improvements to enhance child and family results.

In general, the use of child assessment data to monitor trends or evaluate programs for accountability purposes requires high standards of technical accuracy, particularly if important policy decisions are to be based on their findings. By using sufficient numbers of children and aggregate measures, greater accuracy can be ensured. For preschool children, social indicators, such as the percentage of children in poverty who participate in school readiness programs, are appropriate as aggregate measures. To ensure the greatest validity and reliability, direct assessment of children's learning (cognitive, language, social, or motor) should be conducted by trained examiners under controlled conditions. These types of assessments are costly and burdensome on staff and children, so sampling procedures are often used. Alternatively, program staff can collect valid and reliable child assessment data if the assessment tool is designed for this purpose, if data are not used to make high stakes decisions about individual children, and if staff receive appropriate and ongoing training to conduct the assessments and use the information for curriculum planning that supports the developmental progress of individual and groups of children.

Computerized data entry and management systems are also being developed to aid staff in compiling the information in user-friendly ways to enable aggregation of data in multiple formats—such as displaying child assessment data for a group of children by age, or program quality information by type of program or funding type (e.g., state preschool, general child care, and Head Start). In addition, computerized analysis formats will allow users to generate growth charts over time for individual children and groups of children, or to show graphically programs' strengths or weaknesses with regard to aspects such as program quality variables, child assessment data, and family satisfaction. These potential uses of assessment data are all representative of how program staff and administrators can use the data they generate to inform areas for program improvements that enhance child and family results. These strategies assume that the data collected are directly related to, and broadly representative of, program goals and objectives, are sensitive to cultural and linguistic norms and values and thus appropriate for the populations served, reflect recent theory and research, have practical relevance, and are derived from psychometrically sound measurement strategies (i.e., are valid and reliable).

Training in the use of assessment data for program improvements that enhance child and family results will be needed on an ongoing basis—and as data systems become increasingly available for computerized analysis and reporting, the possibilities will expand for administrators and staff to gain new insights into the strengths and weaknesses of their program. With these new tools at their fingertips, staff will also need assistance in determining how and where to make

adjustments to current practice to improve results for children and families. Regional training and technical assistance networks throughout California are being designed and implemented by the Department of Education, Child Development Division with this purpose in mind, and with particular emphasis on strengthening the integration of assessment results with California's content and performance standards for kindergarten through grade 12.

Analysis of county-level school readiness applications for First 5 California funding revealed that several programs planned to contract with other agencies (e.g. local universities, consultants, or other government agencies) to conduct comprehensive program evaluation. Many programs mentioned they were planning on implementing some type of longitudinal tracking system. However, only a small number of programs mentioned they have school district database systems that allow tracking through 3rd grade. Pre- and post- models were also commonly mentioned. The majority of the applications provided lists of the types of outcomes and indicators to be measured, but did not detail their measurement methods. For example, the majority of the programs mentioned parents' satisfaction with children's progress, outcomes, or quality of care as examples of indicators of program effectiveness. Other indicators included children's successful transition to kindergarten, success in kindergarten, program enrollment, and expansion of the population served. Longer-term indicators mentioned by several programs included improved standardized test scores and decreased numbers of grade retentions.

Uses of Assessment in Program Evaluation

Evaluation of preschool programs to determine whether they are meeting the expected standards of quality and achieving the intended outcomes constitutes a unique purpose for the collection of assessment data. Program evaluations have the potential to showcase promising practices, attract new sources of funding, and inform programmatic and policy decisions. As the stakes increase in terms of program sustainability and public investments, it becomes correspondingly important that the standards for design, instrumentation, and analysis are well-defined and technically rigorous. The assessment methods and instruments for the purpose of program evaluations need to be closely aligned to the program's goals, and special attention must be given to issues of sampling and aggregation of data, to enable appropriate attribution of evaluation findings. Data are gathered about groups of children or families and reported in aggregate form. If sufficient numbers of children or families are sampled, the accuracy of the findings can be ensured. Thus, findings are used to make decisions about programs, and not individuals.

To safeguard against the misuse of child assessment data (e.g., to make decisions about individual children) for the purposes of program evaluation, matrix sampling procedures may be considered.[16] In addition, social or population indicators, that focus on the adequacy of services (e.g., percent of children or families served) or the conditions in the environment (e.g., percent of children living in poverty) are also used as broad snapshots at a single point in time, and to monitor trends over time. Several examples of recent or ongoing program evaluations in California are discussed below.

[16] Matrix sampling is a statistical technique whereby each child takes only part of the total assessment. [See also http://pareonline.net/getvn.asp?v=8&n=16 or http://nces.ed.gov/nationsreportcard/pubs/guide/ques20.asp for more information about matrix sampling]

First 5 California School Readiness Initiative Evaluation and implications for evaluation of preschool programs

The School Readiness (SR) Evaluation underway by SRI, with assistance from AIR, is guided by the First 5 Commission's framework, which is drawn from the National Educational Goals Panel (NEGP) definition of three broad, interrelated components of school readiness:

- Children's readiness for school,
- Schools' readiness for children, and
- Family and community supports and services that contribute to children's readiness for school.

Thus, the school readiness evaluation also lends itself to the three basic types of assessment – child-based, family-based, and program-based, and SRI International has designed the evaluation accordingly. The statewide evaluation includes teacher-administered assessments of entering kindergartners, interviews with families, and surveys of the membership of multiple stakeholder groups (i.e., teachers, County Commission Executive Directors, SR Coordinators, etc.).

The child assessment tool being used is the *Modified Desired Results Developmental Profile* (MDRDP), which incorporates selected measures from the 3 years-through-prekindergarten DRDP and the kindergarten-through-seven years DRDP. The measures were chosen to represent key domains of learning that the SR programs are designed to impact. It is intended to provide a snapshot of children's developmental competencies when they enter kindergarten. Because this single measure is being used for a large sample of children from a statewide representative sample of underperforming schools, it will allow for many statistical comparisons that would not have been possible if different measures were used across the state. Successive cohorts of incoming kindergarteners will be assessed.
The Family Interviews include questions to the parents of incoming kindergarteners about their children's preschool experiences, kindergarten transition experiences, family literacy activities, children's health and motor skills, and family demographics.

Information about the structure, implementation, and quality of the school readiness programs being implemented in each of the sampled counties will also be collected. Together, these data sources will provide a comprehensive picture of the implementation and impact of the First 5-funded school readiness programs across the state.

The statewide evaluation of school readiness programs will provide tremendous insights for participating counties about the impact of their school readiness programs. In addition, counties that are not included in the sample have an opportunity to use a similar design in order to compare themselves to the statewide sample. Through a Memorandum of Understanding between county commissions and the statewide commission, the MDRDP is being made available for use beyond the First 5 SR Evaluation. Counties may wish to use the MDRDP as one of the child measures they can use to obtain a snapshot of entering kindergartners' developmental competencies, so they can monitor local trends over time, in addition to having access to statewide comparison data from the SRI evaluation, as a measure of their relative success. It is important to stress, however, that the MDRDP is not intended as a diagnostic tool

for classifying which children will or will not do well in school. However, it does provide a brief profile of the child's current levels of development in key domains, and is suitable for large groups of children for whom data will be aggregated for the purpose of monitoring trends over time.

School readiness evaluation: Local examples

Many states and localities are interested in identifying indicators of school readiness that can be used to contribute in positive ways to children's early learning and school success. For example, the School Readiness Indicators Initiative works with 17 states to develop a comprehensive set of school readiness indicators to inform public policy for young children and their families. The School Readiness Indicators Initiative has three goals: 1) to create sets of indicators in states to describe school readiness, 2) to use the indicators to influence state policy on children's issues, and 3) to communicate data meaningfully within states and around the country.[17] Regardless of the increasing interest in identifying school readiness indicators and assessment tools, the same cautions with regard to the appropriate uses of assessment in early childhood apply to school readiness assessments. Attention should be given to ethical use of assessment information to benefit children, not harm them. Assessments should be broadly focused, should be embedded in curriculum, should use multiple sources of evidence, and should be based on systematic observation of children over time. In addition, readiness assessments should consider the reciprocal and dynamic role of the learning context and program quality (i.e., the roles of programs, teachers, learning opportunities, and so on) in addressing school readiness issues in the most comprehensive and holistic manner.

A review of county-level proposals for First 5-funded school readiness programs revealed that the most commonly mentioned child assessment strategies included the following: Desired Results Developmental Profiles, "school readiness standards, assessments, or measures" (unspecified), the Brigance, High Scope Child Observation Record (COR), Ages and Stages Questionnaire (ASQ), and portfolios. Many programs mentioned they planned to use some developmental domains and results indicators but did not specify instruments. There were also several that mentioned the use of pre-school exit or kindergarten entry assessments and/or other pre- and post-intervention assessment models. Many programs indicated an emphasis on language and literacy assessments; however, the majority of these programs did not specify the instrument to be used for language assessment. Among the few that did, the Language Assessment Scale (or the Pre-LAS for prekindergarten through first grade) and the California English Language Development Test (administered annually, beginning in kindergarten) were commonly mentioned. Many applications also mentioned longer-term outcome measures such as API scores and SAT 9 scores for following up on children's progress during elementary school.

To assess program quality, school readiness programs most commonly mentioned the ECERS and parent surveys. Many programs also emphasized their attention to the qualifications of teachers, as demonstrated by examining staff resumes and professional records, teachers' knowledge of school readiness standards, and knowledge of child development and program services. Several programs also mentioned "increased articulation/coordination" between

[17] This initiative is sponsored by the David and Lucile Packard Foundation, the Ewing Marion Kauffman Foundation and the Ford Foundation.

prekindergarten and kindergarten staff as a measure of program quality. In general, programs tended to list indicators/outcomes rather than methods of measuring program quality. Examples of indicators and outcomes included "safe and healthy" environment, increased communication between parents and teachers, alignment of curriculum and outcomes, etc. Some programs mentioned the use of data collection systems but did not specify how these were used to assess program quality.

Elk Grove Unified School District's Longitudinal Study is an example of one district's attempt to answer the question, "Does preschool result in strong academic performance in later grades?" EGUSD has been maintaining data on children who were served in the district preschool program since 1989. The district has followed their progress through kindergarten and beyond, comparing their scores on standardized tests (the Stanford Achievement Test, 9th Edition) in later grades with the scores of children from the same high poverty schools who did not attend preschool. This comparison group was used because the demographic characteristics (e.g., ethnicity and socioeconomic status) of these students closely matched those of the children who attended EGUSD preschool programs. Using these data, the district has been able to demonstrate that those children who attended preschool programs located at Title 1 schools scored higher on standardized tests than other students from the same Title 1 schools who did not attend preschool. These results have bolstered EGUSD's efforts to secure additional funding, enabling them to expand their preschool programs to reach a broader number of children. (See additional information about EGUSD preschool programs presented earlier in this report.)

Practical Steps for Communities

Communities wishing to design and implement assessment strategies for use in program improvement and/or evaluation to enhance child and family results have important decisions to consider. Fundamental to the development of a design for local assessment practices is agreement among key stakeholders on the overriding purpose(s) for which an assessment plan is desired. Discussions about the intended purpose(s) of assessment, however, must include representatives from each of the major stakeholder groups that may potentially benefit from the assessment information. In this way, the multiple perspectives of potential end-users are reflected at the onset and consensus can be built and sustained. The following types of questions can guide initial considerations:

- Who are the major stakeholders to benefit from assessment information?
- For what purpose(s) are assessment data required?
- What are the desired results to be achieved for children, families, and programs?
- Who/what are the contributors to these desired results?
- What indicators will demonstrate that the desired results are being achieved?
- At what level(s) (e.g., child, family, program, community, system) will assessment data be collected?
- How can these indicators be measured in ways that are developmentally, culturally, ethnically, and linguistically appropriate?
- Are measurement strategies theoretically sound, technically valid, and reliable sources for information gathering?

- Are proposed measurement tools compatible with the purpose and needs of the community?
- What are the costs of implementing and sustaining assessment plans?
- What safeguards are needed to minimize the risks associated with misuse of assessment data for individual children or families?
- At what level(s) and for whom will assessment data be reported?

Practical steps for developing a system to measure progress and results for programs serving preschool children and their families include the following:[18]

1) Create a vision of the results key stakeholders want for their community.
2) Set and prioritize goals and develop short- and long-term strategies for turning the vision into reality.
3) Decide who is to share responsibility for the assessment design and implementation process. Clarify lines of authority and individual responsibilities.
4) Decide how progress and results will be measured at the individual, family, program, and community levels, as appropriate to the intended program purpose(s). Consider how sharing of processes can help with data collection and information dissemination.
5) Decide how data will be collected and shared at all levels, appropriate to the purpose(s) of assessment. Consider the availability of historical or benchmark data from which to build indicators of progress. Encourage the use of multiple measures so that findings can be corroborated across measures. Link progress and results to fiscal information to determine the unit costs of achieving desired results or to compare the cost-effectiveness of varying approaches to achieving results.
6) Decide which indicators and benchmarks to use, guided by the vision, goals, and intended results of the program. Consider a broad range of possibilities (including existing model indicators currently being used for assessing related national, state, or local child and family initiatives), weigh the costs and benefits of each, and build widespread support for why particular indicators are selected. Early childhood system measurements can be grouped into four categories:
   - What children know and can do
   - Child and family conditions
   - The supply, adequacy, and quality of services
   - Systems capacity
7) Decide how you will know if you have the right benchmarks and indicators
8) Decide how progress and results will be tied to funding or incentives, to promote fiscal accountability and to target available resources to achieve desired results.
9) Decide on what data will be needed, where it will be collected, and how often.
10) Select or develop instruments that are appropriate for measuring the benchmarks and indicators of interest.[19] Consider the validity and reliability of the instruments, their cultural sensitivity to the population(s) to be assessed, burden on respondents in terms of

[18] Adapted from O'Donnell, N. S. and Galinsky, E. (1998) Measuring Progress and Results in Early Childhood System Development. Families and Work Institute. New York.
[19] [ADD references for where to find possible assessment tools. – Burroughs, Mental Measurement, etc.]

time and effort, staff training needs to administer the assessments and interpret the data, and implementation costs (both start-up and ongoing).

11) Decide how to provide financial and technical support to results-based evaluation efforts. Consider the financial resources needed for indicator development, ongoing data collection, and public awareness activities.
12) Foster public "ownership" of the assessment plan by engaging the public in an inclusive process from beginning to end: goal-setting, indicator selection, evaluating progress, and dissemination of findings.
13) Decide how and when goals and procedures can be re-evaluated and revised on a regular basis.

Conclusion

Focusing on the results to be achieved as the starting point for the development of an assessment system helps to provide an overarching target toward which all subsequent decisions and strategies can be directed. This approach ensures that goals are aligned with strategies for assessing progress and that linkages between program inputs, outputs, and initial-, intermediate-, and long-term results are transparent. Nonetheless, the use of assessment for ongoing program improvement to enhance child and family results requires that purposes are clearly defined and agreed upon by key stakeholders from the outset and that rigorous standards for responsible use of data are maintained. For programs serving preschool children and their families, assessment of child progress, family experiences and satisfaction, and program quality together contribute to a rich portrait of the program's achievement of desired results as a whole. Program staff can use multiple sources of information about how they are doing to corroborate findings and target resources to where they are needed most. Community members also can benefit from assessment information by understanding the results of their investments so that scarce resources are allocated to where they can have the greatest impact. Finally, policymakers and community members should keep in mind that short-term program impacts may represent only a fraction of the benefits that can accrue over time. High quality preschool programs that include mechanisms for using assessment data for ongoing program improvements that enhance child and family results will likely show the greatest benefits and cost-savings over the long-term.

References

Abbott-Shim, M. & Sibley, A. (1998*). Assessment Profile for Early Childhood Programs: Research Edition II*. Atlanta, GA: Quality Assist, Inc.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*(4), 541-552.

Bailey, D. B., Jr., McWilliam, R. A., Darkes, L. A., Hebbeler, K., Simeonsson, R. J., Spiker, D., & Wagner, M. (1998, Spring). Family outcomes in early intervention: A framework for program evaluation and efficacy research. *Exceptional Children*, *64*(3), 313-316.

Cross T., Bazron, B., Dennis, K., & Isaacs, M. (1989). Towards a Culturally Competent System of Care, Volume I. Washington, D.C.: Georgetown University Child Development Center, CASSP Technical Assistance Center.

Dunkle, M. & Vismara, L., (2003). A Different Kind of Test, Good Developmental Checkups are Vital for School Readiness, *Education Week,* September 24, 2003, p. 38.

Elk Grove Unified School District (2003). Longitudinal Study of Preschool Student Achievement in the Elk Grove Unified School District – 1989-2002. Report prepared by Nancy Herota, Program Administrator, Learning Support Services – Pre-Kindergarten Programs, Elk Grove, California.

Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Age 3 Through 8. NAEYC. 1990. Available at http://www.naeyc.org/resources/position_statements/pscag98.pdf)

Harms, T., Clifford, R. M., & Cryer, D. (2000). Training session. Assessing Environments with the Early Childhood Environmental Rating Scale-Revised, Infant/Toddler Environment Rating Scale, School-Age Care Environment Rating Scale, and Family Day Care Environment Rating Scale. University of Phoenix, August 16-18, 2000.

Harms, T., Clifford, R., & Cryer, D. (2000) Early Childhood Environment Rating Scale-Revised Edition. New York: Teachers College, Columbia University.

Henderson, A. T., & Berla, N. (1994). A new generation of evidence: The family is critical to student achievement. St. Louis, MO: Danforth Foundation and Flint, MI: Mott (C. S.) Foundation.

Horton, C. and Bowman, B. T. (2002). Child Assessment at the Preprimary Level: Expert opinion and state trends. Herr Research Center, Erikson Institute. Occasional Paper, Number 3.

Jones, Jacqueline. (2003). Early Literacy Assessment Systems: Essential Elements. Educational Testing Service. Princeton, New Jersey. Available on-line: http://www.ets.org/research/pic

Lund, I., Rous, B., Moffett, C.L., Wood, K., & O'Keefe, H. (2002). *State Profiles on Pre-Kindergarten Standards and Accountability Systems*. Lexington, KY: University of Kentucky, State Preschool Accountability Research Collaborative (SPARC). Available on-line: http://www.ihdi.uky.edu/Sparc/states/Intro.pdf

McNaughton, D. (1994). Measuring Parent Satisfaction With Early Childhood Intervention Programs: Current Practice, Problems, and Future Perspectives. *Topics in Early Childhood Special Education,* 14, 26-48.

Meisels, S.J. (1998). Assessing Readiness. National Center for Early Development and Learning, Spotlight No. 3.

Meisels, S.J. & Atkins-Burnett, S. (2000). The Elements of Early Childhood Assessment. In J.P. Shonkoff & S.J. Meisels (Eds.), *Handbook of Early Childhood Intervention.* Cambridge: Cambridge University Press.

Muenchow, S. (2003). A Risk Management Approach to Readiness Assessment. In Scott-Little, C., Kagan, S.L., & Clifford, R.M. (Eds.). *Assessing the State of State Assessment: Perspectives on Assessing Young Children*. Greensboro, N.C.: The Regional Educational Laboratory at SERVE, Associated with the School of Education, University of North Carolina at Greensboro.

The National Association for the Education of Young Children (NAEYC) & The National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2002). *Early Learning Standards: Creating the Conditions for Success.* Washington, DC: Author.

The National Association for the Education of Young Children (NAEYC) & The National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2001). *Still Unacceptable Trends in Kindergarten Entry and Placement.* Washington, DC: Author. Available on-line: http://www.naeyc.org/resources/position_statements/psunacc.htm

The National Association for the Education of Young Children (NAEYC) & The National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003). *Early Childhood Curriculum, Child Assessment, and Program Evaluation: Building an Accountable and Effective System for Children Birth Through Age Eight.* Washington, DC: Author. Available on-line: http://www.naeyc.org/resources/position_statements/Position_Statement_2003.pdf

National Education Goals Panel (1995). *Reconsidering Children's Early Development and Learning: Toward Common Views and Vocabulary.* Washington, DC: Author. Available on-line: http://negp.gov/Reports/child-ea.htm

O'Donnell, N. S. and Galinsky, E. (1998) Measuring Progress and Results in Early Childhood System Development. Families and Work Institute. New York.

Rous, B.S., Moffett, C. L., & Lund, I. A. (2003). *Building Preschool Accountability Systems: Guiding Principles Resulting from a National Study, Technical Report #3.* Lexington: University of Kentucky, Interdisciplinary Human Development Institute, State Preschool Accountability Research Collaborative. Available on-line: http://www.ihdi.uky.edu/Sparc/

Scott-Little, C., Kagan, S.L., & Frelow, V.S. (2003). *Standards for Preschool Children's Learning and Development: Who Has Standards, How Were They Developed, and How Are They Used? Executive Summary*. Greensboro, NC: University of North Carolina, SERVE.

SERVE. (2003). "Institute for Readiness Assessment System: Making Them Work for All Children." Policy Brief. Available on-line: htpp://www.serve.org/Syc/policy.htm

Shepard, L., Kagan, S. L., & Wurtz, E. (eds.). (1998). Principles and Recommendations for Early Childhood Assessments. Washington DC: National Education Goals Panel.

Shepard, L., Kagan, S. L., & Wurtz, E. (2001). Principles and Recommendations for Early Childhood Assessments. The State Education Standard, 2(2), 5-12.

Simeonsson, R.J. (1988). Evaluating the Effects of Famly-Focused Intervention. In D. Bailey and R. Simeonsson (Eds.), *Family Assessment in Early Intervention* (pp. 251-267). Columbus. OH: Charles E. Merrill.

Smith, M.W. and Dickinson, D. K., with Sangeorge, A. and Anastasopoulos. L. (2002). Early Language and Literacy Classroom Observation Toolkit, Research Edition. Baltimore, MD: Brookes Publishing.

Superintendent's Universal Preschool Task Force. (1998). *Ready to Learn: Quality Preschools for California in the 21st Century.* Sacramento: California Department of Education.

U. S. Department of Health and Human Services. (2002, June). *Early Childhood Education and School Readiness: Conceptual Models, Constructs, and Measures* (Workshop Summary, June 17-18, 2002). Washington, DC: Author. Available on-line: http://www.nichd.nih.gov/crmc/cdb/Kyle-workshop.pdf

Whitebook, M., Sakai, L., Howes, C., and Wishard, A. (2003). Accreditation as a Strategy for Improving and Sustaining Quality: A Longitudinal Assessment. Early Childhood Research Quarterly. Vol. 18, No 4.